Lecture 3

Smoothing and optimizing max-relative entropy

Recall the definition of the max-relative entropy, which was introduced in the previous lecture:

$$D_{\max}(\rho \| Q) = \inf\{\lambda \in \mathbb{R} : \rho \leq 2^{\lambda} Q\}. \tag{3.1}$$

In this lecture we will consider what happens when we minimize this function over various choices of ρ and Q. Two common situations in which this is done are as follows:

1. *Smoothing.* For a given ρ, Q, and $\varepsilon > 0$, the *smoothed max-relative entropy* of ρ with respect to Q is defined as

$$D^{\varepsilon}_{\max}(\rho \| Q) = \inf_{\xi \in \mathcal{B}_{\varepsilon}(\rho)} D_{\max}(\xi \| Q), \tag{3.2}$$

   where $\mathcal{B}_{\varepsilon}(\rho)$ denotes the set of states that are ε-close to ρ with respect to some notion of distance.

2. *Optimizing over models.* For a given ρ and a convex set $\mathcal{C}$ of possible choices of models Q, we may consider the quantity

$$D_{\max}(\rho \| \mathcal{C}) = \inf_{Q \in \mathcal{C}} D_{\max}(\rho \| Q), \tag{3.3}$$

   which measures in a certain sense which $Q \in \mathcal{C}$ incurs the least loss of efficiency when serving as a model for the state ρ.

In both cases, the optimizations can be represented as conic programs. We will consider the two types of optimizations separately in the sections that follow.

3.1 Smoothed max-relative entropy

Let us begin with the smoothed max-relative entropy, where one takes the minimum value of the max-relative entropy over all choices of actual states that are close to a given state, as suggested above. The idea is that the smoothed max-relative entropy reflects a tolerance for small errors, which we often have or would like to express when analyzing operationally defined notions. Without smoothing, the max-relative entropy can sometimes, in certain settings at least, have unwanted hyper-sensitivities that smoothing eliminates.

Definition

As it turns out, there is not a single agreed upon definition for the smoothed max-relative entropy; different authors sometimes choose different notions of distance with respect to which the smoothing is done, which translates to different choices for the set $\mathcal{B}_\varepsilon(\rho)$ in (3.2). In addition, the operator ξ is sometimes allowed to range not only over density operators, but also over sub-normalized density operators, and in this case the definition of the max-relative entropy is extended in the most straightforward way to accommodate such operators.

It is typically the case, however, that the notions of distance with respect to which the smoothed max-relative entropy is defined are based on either the trace distance or the fidelity function. Through the Fuchs–van de Graaf inequalities, one finds that the resulting definitions of smoothed max-relative entropy are roughly equivalent, and are certainly quite similar in a qualitative sense.

For the sake of concreteness, we will define the smoothed max-relative entropy in terms of the trace distance, as the following definition makes precise.

Definition 3.1 (Smoothed max-relative entropy)**.** For a density operator $\rho \in \mathrm{D}(\mathcal{X})$, a positive semidefinite operator $Q \in \mathrm{Pos}(\mathcal{X})$, and a real number $\varepsilon \in (0,1)$, the *ε-smoothed relative max-entropy* of ρ with respect to Q is defined as

$$\mathrm{D}^{\varepsilon}_{\max}(\rho\|Q) = \min_{\xi\in\mathcal{B}_\varepsilon(\rho)} \mathrm{D}_{\max}(\xi\|Q), \tag{3.4}$$

where

$$\mathcal{B}_\varepsilon(\rho) = \bigl\{\xi \in \mathrm{D}(\mathcal{X}) \,:\, \tfrac{1}{2}\|\rho - \xi\|_1 \leq \varepsilon\bigr\}. \tag{3.5}$$

A couple of other common choices for $\mathcal{B}_\varepsilon(\rho)$ are these:

$$\begin{aligned} \mathcal{B}_\varepsilon(\rho) &= \bigl\{\xi \in \mathrm{D}(\mathcal{X}) \,:\, \mathrm{F}(\xi,\rho)^2 \geq 1-\varepsilon\bigr\},\\ \mathcal{B}_\varepsilon(\rho) &= \bigl\{\xi \in \mathrm{Pos}(\mathcal{X}) \,:\, \mathrm{F}(\xi,\rho)^2 \geq 1-\varepsilon^2,\ \mathrm{Tr}(\xi) \leq 1\bigr\}. \end{aligned} \tag{3.6}$$

Optimizing over arbitrary closed and convex sets of states

To better understand the smoothed max-relative entropy, it is helpful to consider a more general set-up. Suppose that $\mathcal{C} \subseteq \mathrm{D}(\mathcal{X})$ is any closed and compact set of density operators, let $Q \in \mathrm{Pos}(\mathcal{X})$ be given, and consider the problem of minimizing η over all choices of $\xi \in \mathcal{C}$ and $\eta \in \mathbb{R}$ that satisfy $\xi \leq \eta Q$. This problem can be expressed as a conic problem as will now be described.

First, define a set $\mathcal{K} \subset \mathbb{R} \oplus \mathrm{Herm}(\mathcal{X}) \oplus \mathbb{R} \oplus \mathrm{Herm}(\mathcal{X})$ as follows:

$$\mathcal{K} = \big\{(\lambda, \lambda\xi, \eta, P) \,:\, \eta, \lambda \geq 0,\, \xi \in \mathcal{C},\, P \in \mathrm{Pos}(\mathcal{X})\big\}. \tag{3.7}$$

This set is a closed and convex cone. One may think of $\mathcal{K}$ as being the Cartesian product of three sets: the first is

$$\mathcal{L} = \big\{(\lambda, \lambda\xi) \,:\, \lambda \geq 0,\, \xi \in \mathcal{C}\big\}, \tag{3.8}$$

the second is the set of all nonnegative real numbers $\eta \geq 0$, and the third is $\mathrm{Pos}(\mathcal{X})$. All three of these sets are closed and convex cones; in the case of $\mathcal{L}$ this follows from the assumption that $\mathcal{C}$ is a compact and convex set. It should be noted that the construction of a convex cone $\mathcal{L}$ from a convex set $\mathcal{C}$ like this is both common and useful.

Now, the optimization problem suggested above is evidently equivalent to the problem of minimizing the inner-product

$$\big\langle (0,0,1,0), (\lambda, \lambda\xi, \eta, P)\big\rangle \tag{3.9}$$

over all $(\lambda, \lambda\xi, \eta, P) \in \mathcal{K}$, subject to the affine linear constraints that

$$\lambda = 1 \quad \text{and} \quad \eta Q = \lambda\xi + P. \tag{3.10}$$

By defining a linear map $\phi : \mathbb{R} \oplus \mathrm{Herm}(\mathcal{X}) \oplus \mathbb{R} \oplus \mathrm{Herm}(\mathcal{X}) \rightarrow \mathbb{R} \oplus \mathrm{Herm}(\mathcal{X})$ as

$$\phi(\lambda, X, \eta, Y) = (\lambda, \eta Q - X - Y), \tag{3.11}$$

these affine linear constraints may be expressed as

$$\phi(\lambda, \lambda\xi, \eta, P) = (1, 0). \tag{3.12}$$

The optimization problem being considered is therefore the primal form of a conic program, which is stated below (together with a simplified expression of its dual form) as Optimization Problem 3.2.

The dual form of this optimization problem is given by the maximization of the objective function

$$\big\langle (1,0), (\mu, Z)\big\rangle \tag{3.13}$$

subject to the constraint

$$(0, 0, 1, 0) - \phi^*(\mu, Z) \in \mathcal{K}^*, \tag{3.14}$$

where (μ, Z) ranges over the space $\mathbb{R} \oplus \text{Herm}(\mathcal{X})$ (as is dictated by Optimization Problem 1.4 in Lecture 1). To simplify this problem, we must compute the adjoint map ϕ^* and try to understand what $\mathcal{K}^*$ looks like. The adjoint of ϕ may simply be computed, and one obtains

$$\phi^*(\mu, Z) = (\mu, -Z, \langle Q, Z\rangle, -Z). \tag{3.15}$$

As for the dual cone to $\mathcal{K}$, an element (δ, H, γ, R) is contained in $\mathcal{K}^*$ if and only if

$$\big\langle (\delta, H, \gamma, R), (\lambda, \lambda\xi, \eta, P)\big\rangle = \delta\lambda + \lambda\langle H, \xi\rangle + \gamma\eta + \langle R, P\rangle \geq 0 \tag{3.16}$$

for all $\lambda, \eta \geq 0$, $\xi \in \mathcal{C}$, and $P \in \text{Pos}(\mathcal{X})$. This is equivalent to the requirement that $\gamma \geq 0$, $R \in \text{Pos}(\mathcal{X})$, and

$$\delta \geq -\langle H, \xi\rangle \tag{3.17}$$

for all $\xi \in \mathcal{C}$. By defining a function

$$\psi_{\mathcal{C}}(H) = \inf_{\xi \in \mathcal{C}} \langle \xi, H\rangle, \tag{3.18}$$

we may alternatively express $\mathcal{K}^*$ as follows:

$$\mathcal{K}^* = \big\{(\delta, H, \gamma, R) \,:\, \delta \geq -\psi_{\mathcal{C}}(H),\ \gamma \geq 0,\ R \in \text{Pos}(\mathcal{X})\big\}. \tag{3.19}$$

The dual problem is therefore a maximization of δ subject to the constraint that $\delta \leq \psi_{\mathcal{C}}(Z)$, $\langle Q, Z\rangle \leq 1$, and $Z \in \text{Pos}(\mathcal{X})$, or, equivalently, a maximization of $\psi_{\mathcal{C}}(Z)$ over all $Z \in \text{Pos}(\mathcal{Z})$ satisfying $\langle Q, Z\rangle \leq 1$. We obtain the following simplified expression of this conic program.

Optimization Problem 3.2

Primal problem		*Dual problem*	
minimize:	η	maximize:	$\inf\{\langle \xi, Z\rangle \,:\, \xi \in \mathcal{C}\}$
subject to:	$\xi \leq \eta Q$,	subject to:	$\langle Q, Z\rangle \leq 1$,
	$\xi \in \mathcal{C}$,		$Z \in \text{Pos}(\mathcal{X})$.
	$\eta \geq 0$.		

It can be shown, through the use of Slater's theorem, that strong duality always holds for Optimization Problem 3.2.

Conic program for smoothed max-relative entropy

At this point, one may substitute the set $\mathcal{B}_\varepsilon(\rho)$ for $\mathcal{C}$ in Optimization Problem 3.2 to obtain a conic program for the smoothed max-relative entropy. To be precise, the smoothed max-relative entropy $\mathrm{D}^{\varepsilon}_{\max}(\rho\|Q)$ is the logarithm of the optimal value of this conic program. One may also note that if $\mathcal{C} = \{\rho\}$, then the semidefinite program for the ordinary (non-smoothed) max-relative entropy is recovered.

Naturally, other notions of smoothing can be considered by making alternative choices for the set $\mathcal{C}$.

3.2 Minimizing over a convex set of models

Next we may consider what happens when we minimize the max-relative entropy over a compact and convex set in the second coordinate. That is, for $\mathcal{C} \subseteq \mathrm{Pos}(\mathcal{X})$ being a compact and convex set, we may consider the quantity

$$\mathrm{D}_{\max}(\rho\|\mathcal{C}) = \inf_{Q\in\mathcal{C}} \mathrm{D}_{\max}(\rho\|Q) \tag{3.20}$$

for a given choice of $\rho \in \mathrm{D}(\mathcal{X})$.

Conic program formulation

Going through a similar process to the one above, we obtain the following conic program.

Optimization Problem 3.3

Primal problem

$$\begin{aligned} \text{minimize:}\quad & \eta \\ \text{subject to:}\quad & \rho \leq \eta Q \\ & Q \in \mathcal{C}, \\ & \eta \geq 0 \end{aligned}$$

Dual problem

$$\begin{aligned} \text{maximize:}\quad & \langle \rho, Z\rangle \\ \text{subject to:}\quad & \sup\{\langle Q, Z\rangle \,:\, Q\in\mathcal{C}\} \leq 1 \\ & Z \in \mathrm{Pos}(\mathcal{X}) \end{aligned}$$

Similar to before, the value $\mathrm{D}_{\max}(\rho\|\mathcal{C})$ defined above is the logarithm of the optimal value of this conic program. The constraint in the dual problem can alternatively be written

$$\theta_{\mathcal{C}}(Z) \leq 1, \tag{3.21}$$

where

$$\theta_{\mathcal{C}}(H) = \sup_{Q\in\mathcal{C}} \langle Q, H\rangle \tag{3.22}$$

is the so-called *support function* of the convex set $\mathcal{C}$.

Example 3.4. We have, in fact, already seen an example in which we minimize over models drawn from a convex set: the *conditional min-entropy*. The conditional min-entropy of X given Y, for the state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, is given by

$$\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})_\rho = -\inf_{\sigma \in \mathrm{D}(\mathcal{Y})} \mathrm{D}_{\max}(\rho \| \mathbb{1}_{\mathcal{X}} \otimes \sigma) = -\inf_{Q \in \mathcal{C}} \mathrm{D}_{\max}(\rho \| Q) \tag{3.23}$$

for $\mathcal{C} = \{\mathbb{1}_{\mathcal{X}} \otimes \sigma : \sigma \in \mathrm{D}(\mathcal{Y})\}$. Observe that

$$\begin{aligned} \theta_{\mathcal{C}}(Z) &= \sup_{Q \in \mathcal{C}} \langle Q, Z \rangle = \sup_{\sigma \in \mathrm{D}(\mathcal{Y})} \langle \mathbb{1}_{\mathcal{X}} \otimes \sigma, Z \rangle \\ &= \sup_{\sigma \in \mathrm{D}(\mathcal{Y})} \langle \sigma, \mathrm{Tr}_{\mathcal{X}}(Z) \rangle = \lambda_1\big(\mathrm{Tr}_{\mathcal{X}}(Z)\big). \end{aligned} \tag{3.24}$$

By the dual formulation of the conic program above, we find that

$$\begin{aligned} \mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})_\rho &= \sup\big\{\log\langle \rho, Z \rangle : \lambda_1\big(\mathrm{Tr}_{\mathcal{X}}(Z)\big) \leq 1,\, Z \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})\big\} \\ &= \sup\big\{\log\langle \rho, Z \rangle : \mathrm{Tr}_{\mathcal{X}}(Z) \leq \mathbb{1}_{\mathcal{Y}},\, Z \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})\big\} \\ &= \sup\big\{\log\langle \rho, Z \rangle : \mathrm{Tr}_{\mathcal{X}}(Z) = \mathbb{1}_{\mathcal{Y}},\, Z \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})\big\}, \end{aligned} \tag{3.25}$$

which is consistent with the primal problem in our semidefinite program for the conditional min-entropy from Lecture 2.

Divergence from a convex set of states

For a convex set of states $\mathcal{C} \subseteq \mathrm{D}(\mathcal{X})$, it is typical that one views the quantity

$$\mathrm{D}(\rho \| \mathcal{C}) = \inf_{\sigma \in \mathcal{C}} \mathrm{D}(\rho \| \sigma) \tag{3.26}$$

(where it should be stressed that it is the ordinary quantum relative entropy, not the max-relative entropy, that appears in this equation) as a measure of distance (or divergence) of ρ from $\mathcal{C}$. For example, the *relative entropy of entanglement* of a state $\rho \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$ is given by

$$\mathrm{REE}(\mathsf{Y} : \mathsf{Z})_\rho = \mathrm{D}(\rho \| \mathrm{SepD}(\mathcal{Y} : \mathcal{Z})) = \inf_{\sigma \in \mathrm{SepD}(\mathcal{Y}:\mathcal{Z})} \mathrm{D}(\rho \| \sigma). \tag{3.27}$$

We may consider a similar notion for the *max-relative entropy* in place of the ordinary relative entropy:

$$\mathrm{D}_{\max}(\rho \| \mathcal{C}) = \inf_{\sigma \in \mathcal{C}} \mathrm{D}_{\max}(\rho \| \sigma). \tag{3.28}$$

To better understand this quantity, let us expand the definition of the max-relative entropy, so that we obtain

$$\mathrm{D}_{\max}(\rho \| \mathcal{C}) = \inf\big\{\lambda \in \mathbb{R} : \rho \leq 2^\lambda \sigma,\, \sigma \in \mathcal{C}\big\}. \tag{3.29}$$

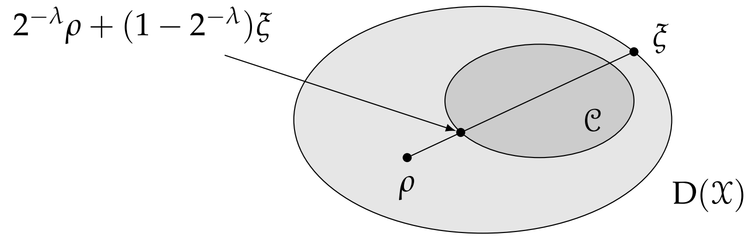

Figure 3.1: $\mathrm{D}_{\max}(\rho\|\mathcal{C})$ is the minimum value of λ for which $2^{-\lambda}\rho + (1-2^{-\lambda})\xi$ is in $\mathcal{C}$, for some choice of a density operator ξ.

The inequality on the right-hand side of this equation can be expressed as an equality through the introduction of a positive semidefinite slack variable P, which yields

$$\mathrm{D}_{\max}(\rho\|\mathcal{C}) = \inf\{\lambda \in \mathbb{R} : \rho + P = 2^{\lambda}\sigma,\, \sigma \in \mathcal{C},\, P \in \mathrm{Pos}(\mathcal{X})\}. \tag{3.30}$$

As ρ and σ are density operators, the slack variable P must have trace equal to $2^{\lambda} - 1$ in order for the equality $\rho + P = 2^{\lambda}\sigma$ to hold. We may therefore replace P by $(2^{\lambda}-1)\xi$ for a density operator ξ, and we obtain

$$\mathrm{D}_{\max}(\rho\|\mathcal{C}) = \inf\{\lambda \in \mathbb{R} : \rho + \big(2^{\lambda}-1\big)\xi = 2^{\lambda}\sigma,\, \sigma \in \mathcal{C},\, \xi \in \mathrm{D}(\mathcal{X})\}, \tag{3.31}$$

which is equivalent to

$$\mathrm{D}_{\max}(\rho\|\mathcal{C}) = \inf\{\lambda \in \mathbb{R} : 2^{-\lambda}\rho + \big(1-2^{-\lambda}\big)\xi \in \mathcal{C},\, \xi \in \mathrm{D}(\mathcal{X})\}. \tag{3.32}$$

This expression reveals that the quantity $\mathrm{D}_{\max}(\rho\|\mathcal{C})$ has the simple and intuitive interpretation suggested by Figure 3.1. For $\lambda = \mathrm{D}_{\max}(\rho\|\mathcal{C})$, the quantity $2^{\lambda}-1$ is sometimes called the (global or generalized) *robustness* of ρ with respect to $\mathcal{C}$.

3.3 Minimizing over both states and models

Finally, and very briefly, it should be noted that one can simultaneously minimize over both arguments in the max-relative entropy. That is, if $\mathcal{C} \subseteq \mathrm{D}(\mathcal{X})$ and $\mathcal{D} \subseteq \mathrm{Pos}(\mathcal{X})$ are convex and compact sets, one may consider the quantity

$$\mathrm{D}_{\max}(\mathcal{C}\|\mathcal{D}) = \inf_{\substack{\sigma\in\mathcal{C}\\ Q\in\mathcal{D}}} \mathrm{D}_{\max}(\rho\|Q). \tag{3.33}$$

This value is the logarithm of the optimal value of the following conic program.

Optimization Problem 3.5

Primal problem		*Dual problem*	
minimize:	η	maximize:	$\psi_{\mathcal{C}}(Z)$
subject to:	$\rho \leq \eta Q$	subject to:	$\theta_{\mathcal{D}}(Z) \leq 1$
	$\rho \in \mathcal{C}$		$Z \in \mathrm{Pos}(\mathcal{X})$
	$Q \in \mathcal{D}$		
	$\eta \geq 0$		

where

$$\psi_{\mathcal{C}}(Z) = \inf_{\rho \in \mathcal{C}} \langle \rho, Z \rangle \qquad \text{and} \qquad \theta_{\mathcal{D}}(Z) = \sup_{Q \in \mathcal{D}} \langle Q, Z \rangle. \tag{3.34}$$